Journal of Psycholinguistic Research, Vol. 32, No. 1, January 2003 (© 2003)

Disfluencies Signal Theee, Um, New Information

Jennifer E. Arnold,[1,2] Maria Fagnano,[1] and Michael K. Tanenhaus[1]

Speakers are often disfluent, for example, saying "theee uh candle" instead of "the candle." Production data show that disfluencies occur more often during references to things that are discourse-new, rather than given. An eyetracking experiment shows that this correlation between disfluency and discourse status affects speech comprehension. Subjects viewed scenes containing four objects, including two cohort competitors (e.g., camel, candle), and followed spoken instructions to move the objects. The first instruction established one cohort as discourse-given; the other was discourse-new. The second instruction was either fluent or disfluent, and referred to either the given or new cohort. Fluent instructions led to more initial fixations on the given cohort object (replicating Dahan et al., *2002). By contrast, disfluent instructions resulted in more fixations on the new cohort. This shows that discourse-new information can be accessible under some circumstances. More generally, it suggests that disfluency affects core language comprehension processes.*

KEY WORDS: reference comprehension; disfluency; language processing; information status.

INTRODUCTION

Most people would like to think of themselves as clear, fluent speakers. But the demands of real-time language use often lead speakers to be disfluent. They may pause, repeat themselves, or restart their utterance. Pauses may be filled with "uh" or "um," and some words may occur with elongated pronunciations, like "theee" (/thij/) for the word "the" (Clark & Wasow, 1998; Fox

Thanks to Sarah Brown-Schmidt and Mikhail Masharov for discussion, and to Katherine Crosswhite for generously analyzing the pitch and duration of the stimuli. This research was partially supported by NIH grant HD-27206.

[1] University of Rochester, Rochester, New York 14627.

[2] To whom all correspondence should be addressed: Department of Brain and Cognitive Sciences, Meliora Hall 495, University of Rochester, Rochester, NY 14627. email: jarnold@bes rochester.edu

Tree & Clark, 1997). Estimates suggest that about 6% of language is disfluent (Fox Tree, 1995).

Even though disfluencies are a pervasive feature of spoken language, they are typically not considered a part of core language processing. As a result, they are frequently excluded from psycholinguistic research and models. At best, disfluencies are considered irrelevant to the comprehender's task of extracting meaning from the linguistic input. At worst, they are seen as a source of processing difficulty. In either case, it is often assumed that the primary reason for identifying disfluencies is so they can be ignored in the comprehension of the remaining linguistic input.

In this study we investigate how disfluency affects on-line language comprehension, focusing on reference resolution. One of the most central aspects of language use is referring—comprehenders must frequently identify who or what the speaker is referring to, and do so quickly. How do comprehenders interpret disfluent references like "theee, uh, candle"?

The standard approach to reference resolution suggests that the referents of definite noun phrases and pronouns are identified through a combination of (a) lexical meaning and (b) discourse constraints that make some referents more accessible than others. The entities that are most accessible are usually considered to be those that have been mentioned recently (i.e., "discourse-given" entities), especially those mentioned in a prominent position (Almor, 1999; Arnold *et al.,* 2000a; Clark & Sengul, 1979; Gordon *et al.,* 1993).

Referent accessibility also influences the forms used for referring. Generally, speakers use pronouns and deaccented definite NPs for referring to highly prominent entities and accented definite NPs for information that is given but less prominent (i.e., unfocused[3]) (Ariel, 1990; Arnold, 1998; Brennan, 1995; Givón, 1983; Gundel *et al.,* 1993). Comprehenders are also sensitive to these patterns. Pronouns are interpreted more quickly when they refer to given/focused information, but this bias is not as strong for fuller expressions (and may be reversed; Hudson-D'Zmura & Tranenhaus, 1998; Gordon & Chan, 1995; Gordon *et al.,* 1993; Gordon & Scearce, 1995). Accenting also affects how comprehenders interpret definite NPs (Dahan *et al.,* 2002). In an eyetracking experiment, accented NPs were resolved most quickly when they referred to given but unfocused information, while deaccented NPs were resolved most quickly when they referred to the most highly focused element in the current discourse.

One recurring theme in research on reference resolution is the idea that given information is more accessible to comprehenders than new information. This bias could be explained by the "expectancy hypothesis": Given informa-

[3] In this paper we are using the term *focus* in the sense of "focus of attention," and not in the sense of the information status in contrast with *topic.*

tion is more accessible because it is relatively more likely that the speaker will continue talking about the same thing than about something new (Arnold, 1998, 2001). Even though speakers switch to new topics all the time, the likelihood of any particular new thing being mentioned is very low, compared to all other new things. Therefore, new information usually has very low expectancy. However, the expectancy hypothesis predicts that if new information is likely to be mentioned, it will also have high expectancy. In this case reference resolution for new items should be relatively faster, possibly faster than for given items.

One reason that speakers are disfluent is because they are having difficulty with some aspect of language production (Clark & Wasow, 1998; Fox Tree & Clark, 1997). If reference to new information causes production difficulty, then we should see more disfluency in reference to new than given information. This pattern indeed exists, as shown by an analysis of data collected for a naturalistic production study (Arnold *et al.*, 2000b). In this study, pairs of subjects gave instructions to each other about how to move objects, which were either discourse-given (usually mentioned in the previous utterance) or discourse-new. An analysis of all NPs ($n = 5128$) shows that 21% of new NPs were disfluent, but only 16% of given NPs were disfluent ($\chi^2 = 19.56$, $p < .001$). This tendency for speakers to be disfluent when referring to new objects is also supported by Barr's (2001a,b) production experiment.

Therefore, the expectancy hypothesis suggests that listeners might use disfluencies as a probabilistic cue that the speaker is less likely to be referring to something recently mentioned and more likely to be referring to an entity that is discourse-new or otherwise relatively inaccessible. We investigated this prediction by tracking participants' eye movements as they responded to instructions to move objects on a computer screen. Eye movements are an ideal measure for investigating how comprehenders interpret referring expressions in tasks like these. In order to perform the task of moving objects, subjects usually fixate the target object as soon as they identify it. As the speech signal unfolds over time, eye movements reveal the objects that are being considered as referents for the expression (Tanenhaus *et al.*, 1995, 1996). Eye movements are also fast and unconscious and thus provide a window to listeners' on-line interpretation of fluent and disfluent referring expressions.

The hypothesis that disfluency affects on-line reference comprehension is strengthened by recent research showing that disfluencies do not always hinder language processing. Although false starts impair word monitoring, repetitions do not (Fox Tree, 1995), and the presence of "uh" actually helps (in comparison with "um," which was found to have no effect; Fox Tree, 1999). Disfluent fillers like "uh" have also been found to help comprehenders recover from false information in repairs (Brennan & Schober, 2001). Bailey and Ferreira (2002) present evidence that disfluency can also affect grammaticality judgements of sentences with syntactic ambiguities. They conclude that disfluency serves as a cue to upcoming structure and also affects parsing by prolonging the time that the parser is committed to a particular analysis.

There is also evidence that disfluency affects the comprehension of reference to new and old objects, at least off-line. In Barr's (2001a) reference comprehension experiment, listeners initiated movement of the mouse toward a new object on the computer screen more quickly for expressions with a pause and an "um" than for expressions with an irrelevant noise, (e.g., a cough). However, eye movement data did not corroborate this finding. A second experiment (2001b) showed that "uh" and "um" do not differ from each other when placed in pauses of similar length. By contrast with the current experiment, Barr used novel stimuli with descriptions like "the cake with the chunky candles." The length of the target phrases makes it difficult to observe the time course of disfluency effects and leaves open the question of how disfluent instructions differ from fluent ones.

The current study investigated the on-line effects of disfluency on the comprehension of definite noun phrases, e.g., "theee, uh, candle." Each trial presented participants with four well-known objects, including two cohort competitors with names that had overlapping initial segments (candle, camel). As spoken language unfolds over time, the process of word recognition involves the temporary activation of words that are consistent with the input (Marslen-Wilson, 1987), and listeners tend to look at objects they are considering as potential referents (Allopenna *et al.,* 1998; Tanenhaus *et al.,* 1996). Thus, the expression "candle" should lead to initial fixations on both the candle and the camel.

This "cohort competitor effect" is also sensitive to the relative likelihood of a given object being a referent. For example, listeners tend to look more at objects with frequent names (Dahan *et al.,* 2001) or at the preferred objects for accented or deaccented NPs (Dahan *et al.,* 2002). Thus, this approach allows us to investigate fine-grained effects of disfluency and discourse status on the preference for target and competitor objects.

METHOD

Participants

Twenty-four native English speakers from the University of Rochester community participated in exchange for $7.50.

Materials, Design, and Procedure

Participants followed pairs of instructions to move objects on a computer screen, as in Fig. 1.

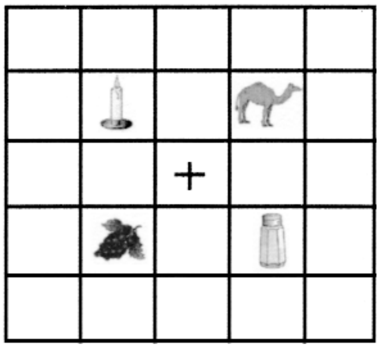

Fig. 1. Sample visual display containing four objects (candle, camel, grapes, salt shaker).

On each trial the scene contained two cohort competitor objects (e.g., candle, camel) and two distractors (e.g., grapes, salt shaker). Given the documented cohort competitor findings (Allopenna *et al.*, 1998; Marslen-Wilson, 1987; Tanenhaus *et al.*, 1996), we expected participants to look at both the target and competitor during the initial, ambiguous portion of the word.

Participants heard prerecorded instructions like those in (1).

(1) a. DISCOURSE-OLD CONTEXT: Put the grapes below the candle.
DISCOURSE-NEW CONTEXT: Put the grapes below the camel.
b. FLUENT: Now put <u>the candle</u> below the salt shaker
DISFLUENT: Now put <u>theee, uh, candle</u> below the salt shaker.

The target noun phrase (underlined in (1b)) was either fluent (the candle) or disfluent (theee, uh, candle). The disfluency involved pronouncing “the” as “theee” (/thij/), inserting the filled pause “uh,” and using a disfluent prosody (lengthened word durations and a disfluent pitch contour). The first sentence established either the target or competitor as given (but unfocused). When the target was given, the competitor was new, and vice versa.

Reference to the target NP was either fluent and accented or disfluent. Fluent and accented NPs preferentially refer to the unfocused second-mentioned entity (Dahan *et al.*, in press). The disfluent NP, by contrast, was hypothesized to make the discourse-new entity more available.

Participants were told that the instructions had been generated by another subject, in the context of the same visual scene, in an earlier phase of the study (but in fact they had been recorded by the first author). It was emphasized that the speaker had been shown **what** to say by graphic cues but had to come up with her own words. This story was necessary to justify the disfluent production and to encourage comprehenders to approach the task in a manner as similar as possible to that used for natural speech situations. A postexperiment questionnaire confirmed that subjects believed the story and generally found the instructions natural.

The 16 experimental items were rotated through these four conditions and combined with 32 filler items. The target item (camel vs. candle) was also manipulated as an additional control variable, resulting in eight lists, with both forward and backward versions. All the filler items contained cohorts; half of them began like the target items but did not mention either cohort in the second utterance. The other half mentioned no cohort in the first utterance. Half the filler instructions contained disfluencies of various types and in various locations. The visual stimuli were versions of the original Snodgrass and Vanderwart (1980) pictures, colored and normed for frequency, visual complexity, and familiarity (Rossion & Purtois, 2001). The frequency, visual complexity, and familiarity of the three cohort words were counterbalanced across items, such that on average these properties were the same for the target and competitor (cf. Dahan *et al.*, 2001). The location of the cohort items was also counterbalanced across items.

Participants' eye movements were recorded with an Applied Scientific Laboratories head-mounted eyetracker. Fixations with respect to the visual scene were recorded on a Hi-8 videorecorder with frame-accurate sound. Eye movements were hand coded, beginning at the onset of the second sentence, in terms of where the subject was looking for each 33 ms frame on the videotape.

PREDICTIONS

The data from the fluent condition were predicted to replicate the findings of Dahan *et al.* (2002) and show more initial fixations on the competitor when it was given and the target was new, in comparison with the condition where the target was given. The condition of interest was the disfluent condition, where there were separate predictions for the region starting at the onset of the determiner (the/ theee) and the region starting at the onset of the head noun. If the disfluency makes discourse-new entities more expected as referents, participants should show more fixations on all discourse-new objects (both the new cohort and the new unrelated) as soon as the disfluency is detected. After the onset of the head noun, this expectancy should translate into more fixations on the new cohort—that is, more fixations on the competitor in the disfluent/given condition, and little competition in the disfluent/new condition.

RESULTS AND DISCUSSION

We examined the percentage of fixations on target, competitor, and unrelated items during two regions: (a) after the onset of the determiner and (b) after the onset of the head noun. In both cases, we analyzed a window of 300 ms, discounting the first 200 ms.[4]

Early Effects of Disfluency

We looked for the early effects of the disfluency in the region 200–500 ms after the onset of the determiner. As predicted, there were more fixations on all new objects (the new cohort and the salt shaker) in the disfluent condition (see Fig. 2). This suggests that disfluency leads comprehenders to immediately focus their attention on new objects. There was an average of 988 ms between the onset of "theee, uh" and the onset of the head noun,

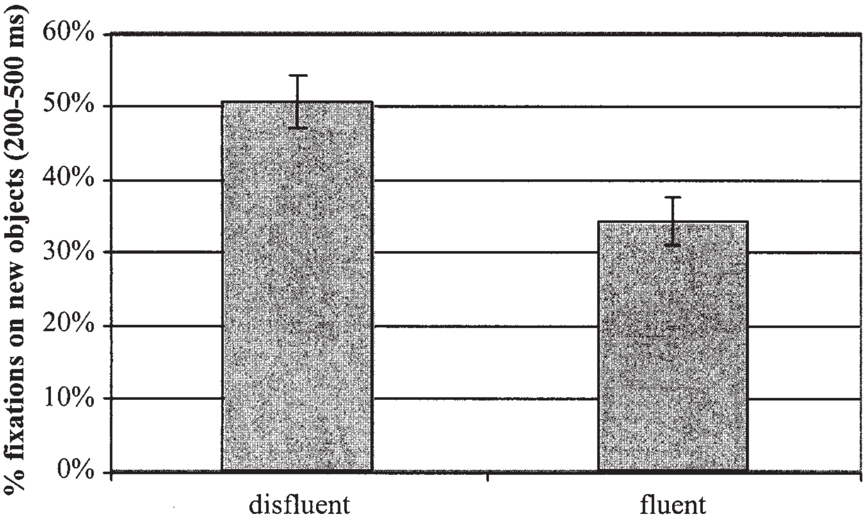

Fig. 2. Percentage of fixations on all new objects from 200 to 500 ms after the onset of "the"/"theee uh."

[4] It takes about 200 ms to program and launch an eye movement in a visual scene with multiple potential targets (Fisher, 1992) so we did not expect signal-driven differences to occur until at least 200 ms after the onset of the input of interest, an estimate that has been consistently confirmed in other research with displays similar to those used here (e.g., Allopenna *et al.,* 1998; Dahan *et al.,* 2001).

which means that this preference for new objects is occurring entirely before the target noun occurs.

By contrast, in the fluent condition there were more fixations on given objects (the given cohort and the grapes) than on new objects. In this condition, there was an average of only 102 ms between the onset of "the" and the onset of the head noun, so the bias toward given objects is mostly carried by an early preference for the given cohort object.

The same pattern occurs if we look only at cohort objects (i.e., target and competitor). In the disfluent conditions there were 28% fixations on the new cohort, and in the fluent condition there were 17%.

Analyses of variance were performed over both subject and item means in the four conditions, with "fixations on new objects" as the dependent variable. The preference for new items in the disfluent condition and given items in the fluent condition is reflected in a main effect of disfluency ($F1(1,23) = 23.05$, $p < .001$; $F2(1,15) = 14.47$, $p < .005$).

Effects on Reference Resolution

It is clear from the first analysis that disfluency leads to an early bias toward new objects. How does this affect reference resolution?

One possibility is that the early bias to new objects combines with a pervasive bias toward given objects during reference resolution. If this is the case, we would expect to see equal looks to both given and new targets in the disfluent condition. A second possibility is that the disfluency introduces an expectancy for reference to new objects that translates into facilitation for the new cohort during reference resolution. In this case, we expect more fixations on the new cohort in the disfluent condition. In both cases, we expect the fluent instructions to lead to more early fixations on the given cohort.

An analysis of the fixations after the head noun shows that there are more fixations on the discourse-new cohort in the disfluent condition, supporting the second interpretation. An analysis of fixations on target and competitor objects reveals a bigger "target advantage" in the disfluent/new and fluent/given condition, meaning that there were more fixations on the target than competitor objects. By contrast, the disfluent/given and fluent/new conditions led to more fixations on the competitor cohort (Fig. 3). Analyses of variance on target fixations minus competitor fixations show the predicted interaction between discourse status and disfluency ($F1(1,23) = 11.01$, $p < .005$, $F2(1,15) = 12.99$, $p < .005$).

Thus, the data in Fig. 3 support the idea that disfluency increases the accessibility of discourse-new objects during reference resolution. One concern with this conclusion, however, is that the data pattern after the head noun might result from the baseline differences between the fluent and disfluent

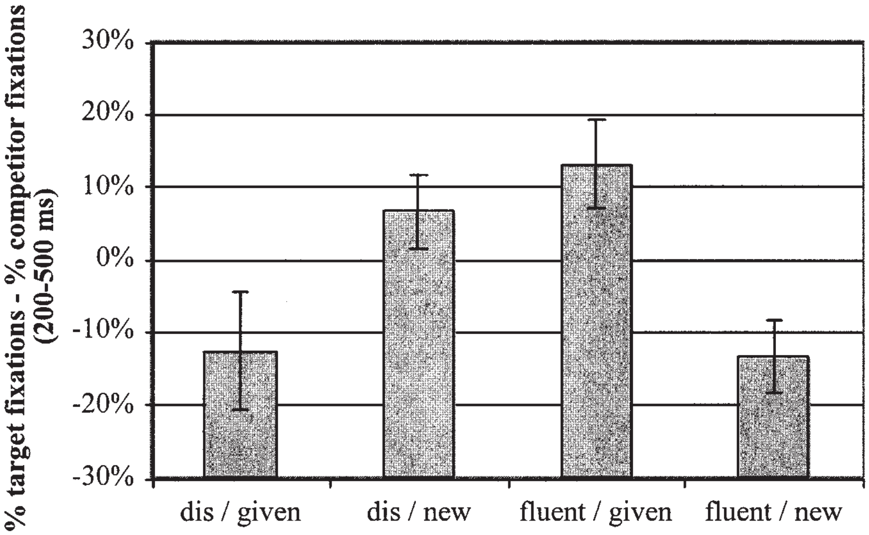

Fig. 3. Percentage of target fixations minus percentage competitor fixations in each condition, 200–500 ms after the onset of the head noun.

conditions. We wanted to know whether the early bias in the disfluent condition translated into facilitation for the new cohort when the head noun was encountered (as opposed to a continued bias toward all discourse-new objects).

We therefore conducted an additional analysis on a subset of the data, including only those trails in which the participant was not already looking at the target or competitor at the onset of the head noun. This eliminates any prior bias to fixate discourse-new objects. After excluding all cases where the target or competitor was fixated at the onset of the head noun, 64% of the data were left.[5]

This subset of data was analyzed from 200 to 500 ms after the onset of the head noun. Figure 4 shows that, again, there is a bigger target advantage in the disfluent/new and fluent/given conditions. This pattern is supported by analyses of variance over subject and item means, with the dependent variable as fixations on the target minus fixations on the competitor, which again shows an interaction between discourse status and disfluency ($F1(1,23) = 4.60$, $p < .05$; $F2(1,15) = 9.35$, $p < .001$).

Thus, there are two patterns emerging in these data. First, the disfluent condition leads to a clear early bias toward new objects. Second, the continued bias toward new objects after the head noun suggests that disfluency facilitates

[5] Three subjects had missing data in one condition; these data were replaced by the subject mean.

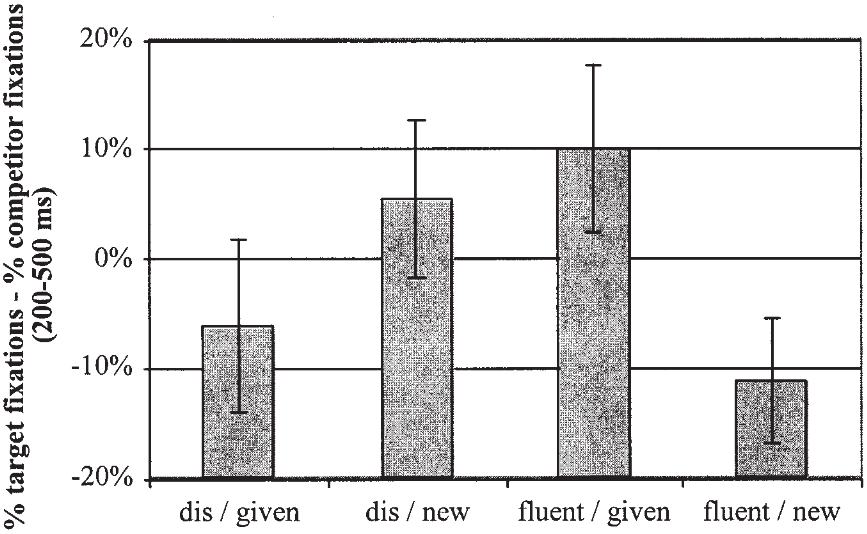

Fig. 4. Percentage of target fixations minus percentage competitor fixations in each condition. Fixations cover 200–500 ms after the onset of the head noun, including only those trials where the subject was not looking at the target or competitor at the onset of the head noun.

the identification of the discourse-new object as referent. This contrasts with the fluent conditions, which lead to a bias for the given cohort object. These effects of disfluency on reference resolution are occuring as early as other documented effects (e.g., Allopenna *et al.*, 1998; Arnold *et al.*, 2000a), suggesting that disfluency is used for on-line language processing.

What are the differences between fluent and disfluent instructions that give rise to this effect? The disfluent instructions in this study included the filler "uh," confirming that the bias toward new objects is not limited to "um" (cf. Barr, 2001a, 2001b). However, this was not the only manifestation of disfluency: The disfluent conditions also contained an elongated "theee" (/thij/), longer word durations—as early as the words *Now* and *put*—and a different pitch contour. The resulting instruction was a natural conglomeration of features that lead to the impression of speaker disfluency, but further research is needed to identify the contribution of each of these features.

CONCLUSION

This study points to two general conclusions about language processing. First, disfluency affects core language comprehension processes. Disfluency creates a bias that an upcoming referring expression is less likely to refer to a just-mentioned referent. Once the lexical information becomes available, the bias toward new information combines with it to facilitate reference to a new object and interfere with reference to a given object.

The second conclusion supported by these data is that discourse-new objects can be more accessible than discourse-given objects in some conditions. This finding challenges the traditional explanation of discourse accessibility. Typically, entities are considered accessible if they have been prominent in the preceding discourse, or at the very least if they are given information. Explanations of accessibility in terms of the history of the discourse, or how an entity has previously been mentioned, do not account for the faster identification of new referents for disfluent referring expressions.

The bias toward new objects instead supports the expectancy hypothesis. Disfluency is a cue that the speaker is probably referring to new information. Like most spoken language use, this experiment presented listeners with a restricted referential domain. This meant that disfluency specifically increased the expectancy of objects that were visible but had not just been mentioned, thus making them more accessible. By contrast, fluent utterances created the familiar bias toward given information. These two findings together can be explained by the idea that accessibility is modulated by expectancy.

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, *106*(4), 748–765.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents.* London: Routledge.

Arnold, J. (1998). *Reference Form and Discourse Patterns.* Ph.D. Dissertation, Stanford University.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference Continuation. *Discourse Processes*, *31*(2), 137–162.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000a). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76,* B13–B26.

Arnold, J., Wasow, T., Ginstrom, R., & Losongco, T. (2000b). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*(1), 28–55.

Bailey, K. G. D., & Ferreira, F. (2002). Disfluencies affect the parsing of garden-path sentences. Manuscript, Michigan State University.

Barr, D. J. (2001a). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.), *Oralité et gestualité: Interactions et comportements multimodaux dans la communication* (pp. 597–600). Paris: L'Harmattan.

Barr, D. J. (2001b). Paralinguistic correlates of discourse structure. Poster presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL, Nov. 15–18.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes, 10*(2), 137–167.

Brennan, S. E., & Schober, M. E. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*(2), 274–296.

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory and Cognition, 7,* 35–41.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37,* 201–242.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language, 47,* 292–314.

Fisher, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and Visual Cognition: Scene Perception and Reading* (pp. 31–45). New York: Springer-Verlag.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34,* 709–738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29*(2), 320–326.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62,* 151–167.

Givón, T. (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study.* Amsterdam: John Benjamins Publishing.

Hudson-D'Zmura, S., & Tanenhaus, M. K. (1998). Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment. In M. Walker, A. Joshi, & E. Prince (Eds.), *Centering Theory in Discourse* (pp. 199–226). Oxford: Oxford University Press.

Gordon, P. C., & Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language, 34*(2), 216–231.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17,* 311–347.

Gordon, P. C., & Scearce, K. A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory and Cognition, 23*(3), 313–323.

Gundel, J. K., Hedberg, N., & Zacharaski, R. (1993). Cognitive status and the form of referring expressions. *Language, 69*(2), 274–307.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25,* 71–102.

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition. Paper presented at the 1st Vision Science Conference, Sarasota, Florida, May, 2001.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 174–215.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In T. Inui & J. L. McClelland (Eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 457–478). Cambridge, MA: MIT Publishing.